


<p><b>Nama:</b> <b>Ibnu Fajar Setiawan</b></p> <p><b>NIM:</b> <b>065002000006</b></p>	 <p><b>Praktikum Data Warehouse</b></p>	<p><b>MODUL 9</b></p> <p><b>Nama Dosen:</b> <b>Ir. Teddy Siswanto, MMSi</b></p>
<p><b>Hari/Tanggal:</b> <b>Hari, 06/06/2022</b></p>		<p><b>Nama Asisten Labratorium:</b>  <b>1. Azhar Rizki Zulma</b>  <b>065001900001</b>  <b>2. Nadiya Amanda Rizkania</b>  <b>064001900003</b></p>

## Proyek Akhir Praktikum I

### 1. Teori Singkat

Data warehouse adalah jenis sistem manajemen data yang dirancang untuk memungkinkan dan mendukung kegiatan business intelligence (BI), terutama analitik. Gudang data semata-mata dimaksudkan untuk melakukan kueri dan analisis dan sering berisi sejumlah besar data historis. Data dalam gudang data biasanya berasal dari berbagai sumber seperti file log aplikasi dan aplikasi transaksi. Gudang data memusatkan dan mengkonsolidasikan sejumlah besar data dari berbagai sumber. Kemampuan analitisnya memungkinkan organisasi untuk memperoleh wawasan bisnis yang berharga dari data mereka untuk meningkatkan pengambilan keputusan. Seiring waktu, ia membangun catatan sejarah yang dapat sangat berharga bagi para ilmuwan data dan analis bisnis. Karena kemampuan ini, gudang data dapat dianggap sebagai "sumber kebenaran tunggal" organisasi.

### Cleansing Data

Cleansing Data adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat. Data tidak akurat tersebut berupa data yang kosong dan tidak memiliki nilai atau berisi nilai NULL/NaN. Tujuan dari cleansing data ialah agar ketika data diproses nanti, tidak terjadi adanya error dari pengolahan data yang dilakukan akibat data yang tidak lengkap ataupun data hilang (missing value).

Berikut ini adalah langkah langkah dalam melakukan cleansing data menggunakan Spoon Pentaho:



## 1. Data Preparation

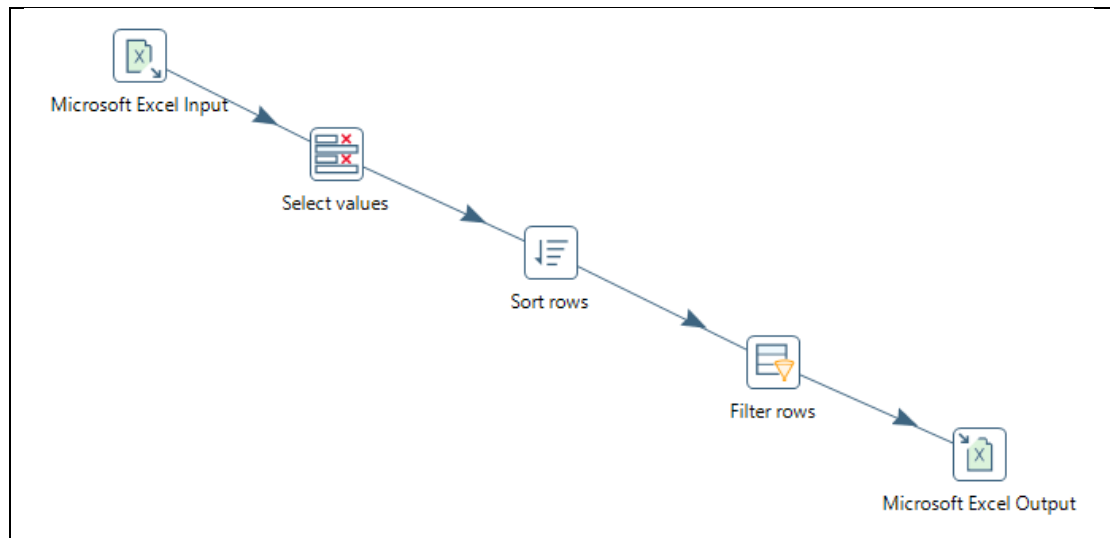
Unduh serta ekstrak datanya dan satukan kedalam folder seperti pada gambar dibawah.



Name	Date modified	Type	Size
Cabai Merah Keriting Desember 2018	14/11/2020 21:16	Microsoft Excel C...	10 KB
Cabai Merah Keriting November 2018	14/11/2020 21:14	Microsoft Excel C...	10 KB
Cabai Merah Keriting Oktober 2018	14/11/2020 21:02	Microsoft Excel C...	10 KB
Cabai Merah Keriting September 2018	14/11/2020 20:45	Microsoft Excel C...	10 KB
data cabai keriting merah 2018 (jan ags)	01/05/2021 12:10	Microsoft Excel 97...	108 KB

## 2. Excel Input

- Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format ***Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr*** seperti ***Cleansing Cabai Merah Keriting Agustus 2018.ktr***



- Berikut ini adalah konfigurasi pada step *Microsoft Excel Input* (Files, Sheets, Fields)

The image displays two screenshots of the 'Microsoft Excel input' configuration window, showing the 'Files' and 'Sheets' tabs.

**Top Screenshot (Files Tab):**

- Step name: Microsoft Excel Input
- Spread sheet type (engine): Excel 97-2003 XLS (JXL)
- File or directory: [Empty field]
- Regular Expression: [Empty field]
- Exclude Regular Expression: [Empty field]
- Password: [Empty field]
- Selected files:
 

#	File/Directory
1	C:\Users\Azhar Rizky Zulma\OneDrive\Documents\Job\Data Warehouse\Praktikum Video Cleanshing and Filter
- Buttons: Add, Browse..., Delete, Edit
- Accept filenames from previous steps:
  - Accept filenames from previous step: ☐
  - Step to read filenames from: [Dropdown]
  - Field in the input to use as filename: [Dropdown]
- Show filename(s)...
- Buttons: Help, OK, Preview rows, Cancel

**Bottom Screenshot (Sheets Tab):**

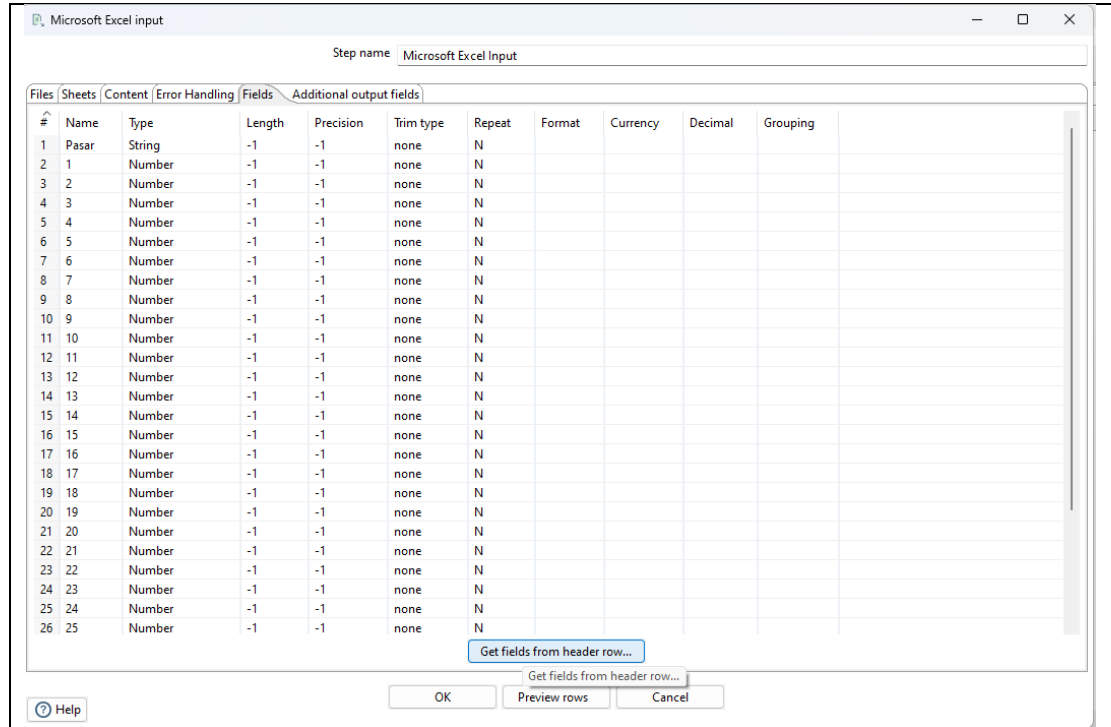
- Step name: Microsoft Excel Input
- List of sheets to read:
 

#	Sheet name	Start row	Start column
1	agustus 2018	0	0
- Buttons: Get sheetname(s)...
- Buttons: Help, OK, Preview rows, Cancel

**Enter list Dialog Box:**

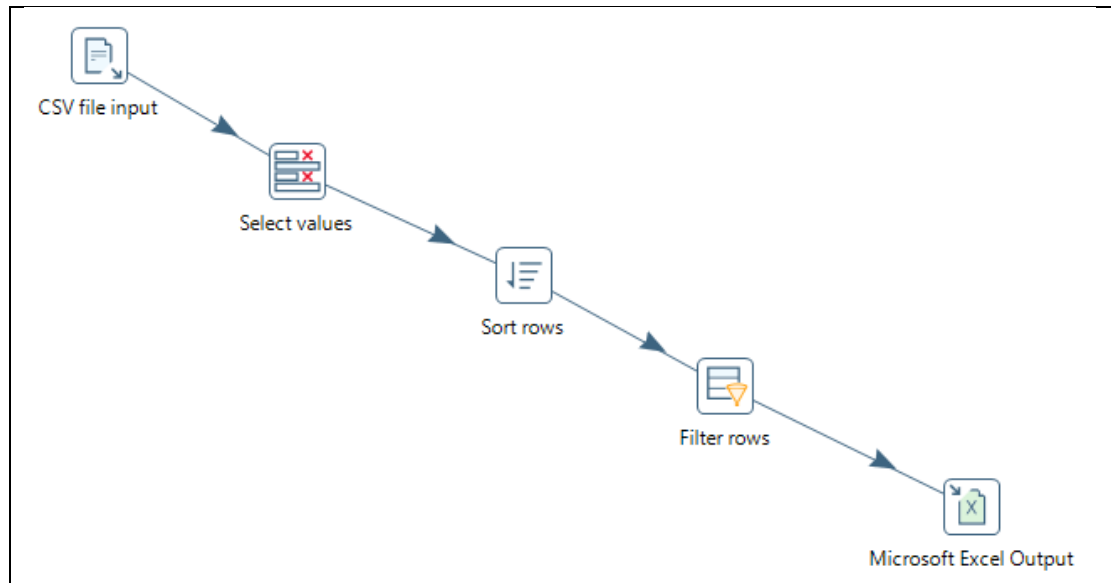
- Available items:
  - januari 2018
  - februari 2018
  - maret 2018
  - april 2018
  - mei 2018
  - juni 2018
  - juli 2018
- Your selection:
  - agustus 2018
- Buttons: Add the selected items on the left, OK, Cancel





### 3. CSV Input

- Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format ***Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr*** seperti ***Cleansing Cabai Merah Keriting Oktober 2018.ktr***



- Berikut ini adalah konfigurasi pada step CSV File Input.

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	No	Integer	#	15	0	IDR	.	,	none
2	Pasar	String		26		IDR	.		none
3	1	Integer	#	15	0	IDR	.	,	none

#### 4. Data Filtering

- Step selanjutnya adalah melakukan select value yang berguna untuk memilah kolom pada excel seperti pada gambar berikut. Kemudian Delete kolom No saja, karena nomor tidak diperlukan lalu Klik OK.

#	Fieldname	Rename to	Length	Precision
1	Pasar			
2	1			
3	2			
4	3			
5	4			
6	5			
7	6			



- Kemudian lakukan ascending data pada proses step Sort Rows

Sort rows

Step name:

Sort directory:

TMP-file prefix:

Sort size (rows in memory):

Free memory threshold (in %):

Compress TMP Files? ☒

Only pass unique rows? (verifies keys only) ☐

Fields:

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	Pasar	Y	N	N	2	N
2	1	Y	N	N	2	N
3	2	Y	N	N	2	N
4	3	Y	N	N	2	N
5	4	Y	N	N	2	N

- Kemudian pada step Filter Rows, buang beberapa nama pasar seperti contoh pada dibawah ini. Untuk menambahkan condition klik icon + di pojok kanan atas.

Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

AND

AND

AND

AND



Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

Level 1. Select UP to go up one level

(String)

- Berikut List Data Nama Pasar yang mau dibuang (2018)

	NOT ( Pasar = [Pasar Anyer Bahari] )
AND	NOT ( Pasar = [Pasar Gondangdia] )
AND	NOT ( Pasar = [Pasar Induk Kramat Jati] )
AND	NOT ( Pasar = [Pasar Kalibaru] )
AND	NOT ( Pasar = [Pasar Kalideres] )
AND	NOT ( Pasar = [Pasar Klender SS] )
AND	NOT ( Pasar = [Pasar Koja Baru] )
AND	NOT ( Pasar = [Pasar Ravanangun] )
AND	NOT ( Pasar = [Pasar Pramuka] )
AND	NOT ( Pasar = [Pasar Pluit] )
AND	NOT ( Pasar = [Pasar Pulo Gadung] )
AND	NOT ( Pasar = [Pasar Pesanggrahan] )
AND	NOT ( Pasar = [Pasar Glodok] )
AND	NOT ( Pasar = [Pasar Matraman KK] )
AND	NOT ( Pasar = [Pasar Jatinegara] )



## 5. Excel Output

Langkah terakhir simpan data menggunakan step *Microsoft Excel Output* dengan konfigurasi sebagai berikut lalu simpan dengan format nama ***Cabai Merah Keriting [Bulan] [Tahun].xls*** seperti ***Cabai Merah Keriting Agustus 2018.xls*** (File & Sheet)

The top screenshot shows the 'Microsoft Excel output' window with the 'File' tab selected. The 'Step name' is 'Microsoft Excel Output'. The 'Filename' is 'C:\Users\Azhar Rizky Zulma\OneDrive'. The 'Extension' is 'xls'. The 'Add filenames to result' checkbox is checked. The 'Date time format' dropdown is empty. The 'Show filename(s)...' button is visible. The 'OK' and 'Cancel' buttons are at the bottom.

The bottom screenshot shows the 'Microsoft Excel output' window with the 'Fields' tab selected. The 'Step name' is 'Microsoft Excel Output'. The 'Fields' tab shows a table with 6 fields:

#	Name	Type	Format
1	Pasar	String	
2	1	Integer	
3	2	Integer	
4	3	Integer	
5	4	Integer	
6	5	Integer	

The 'Get Fields' and 'Minimal width' buttons are at the bottom of the table. The 'OK' and 'Cancel' buttons are at the bottom of the window.





## 2. Alat dan Bahan

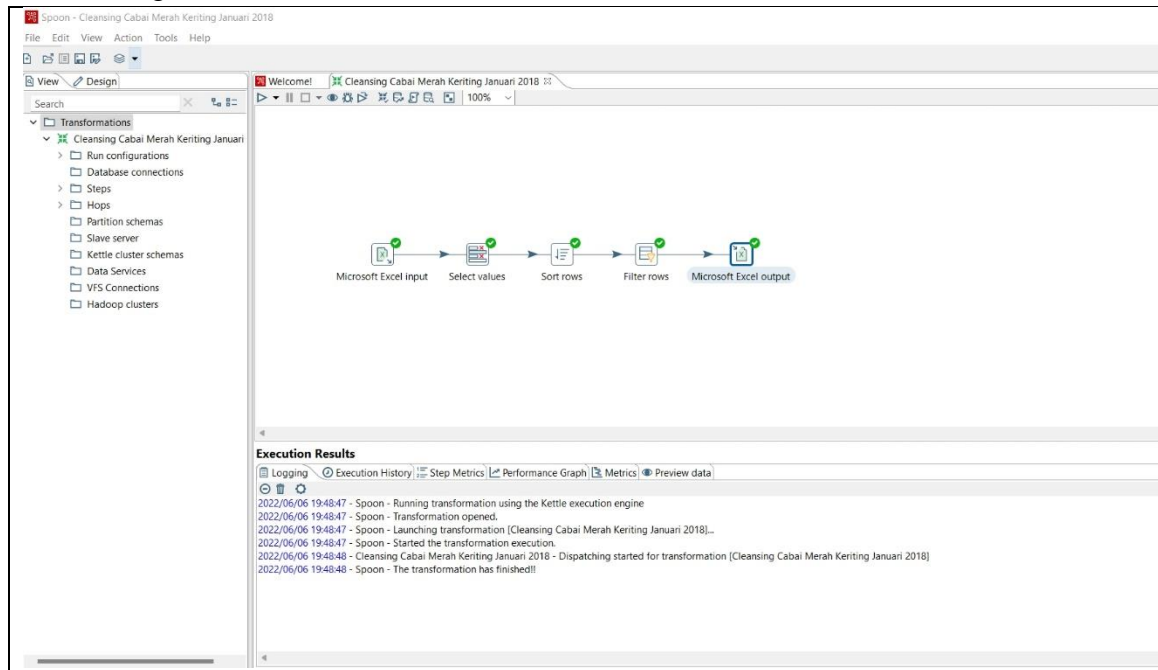
Hardware : Laptop/PC

Software : Spoon Pentaho from Hitachi Vantara

## 3. Elemen Kompetensi

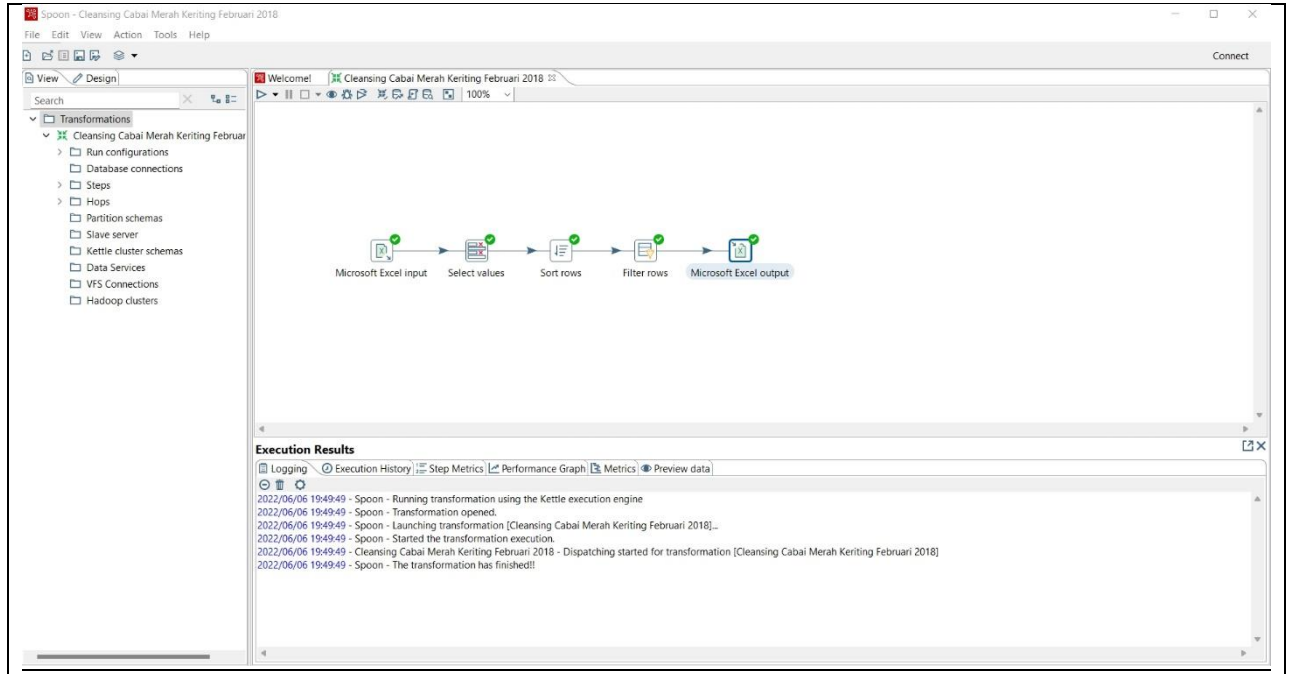
a. Tugas Akhir Praktikum I – Screenshot Struktur yang sudah berhasil berjalan

### 1. Cleansing Januari 2018

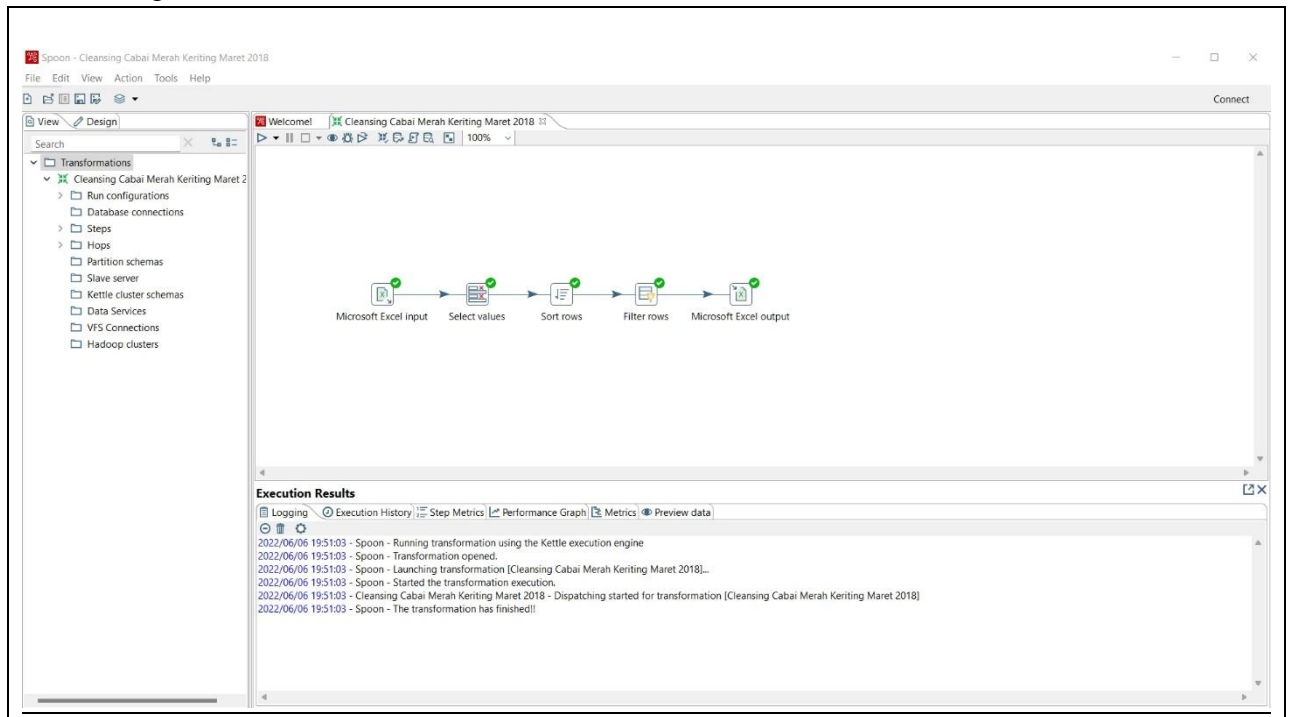


### 2. Cleansing Februari 2018



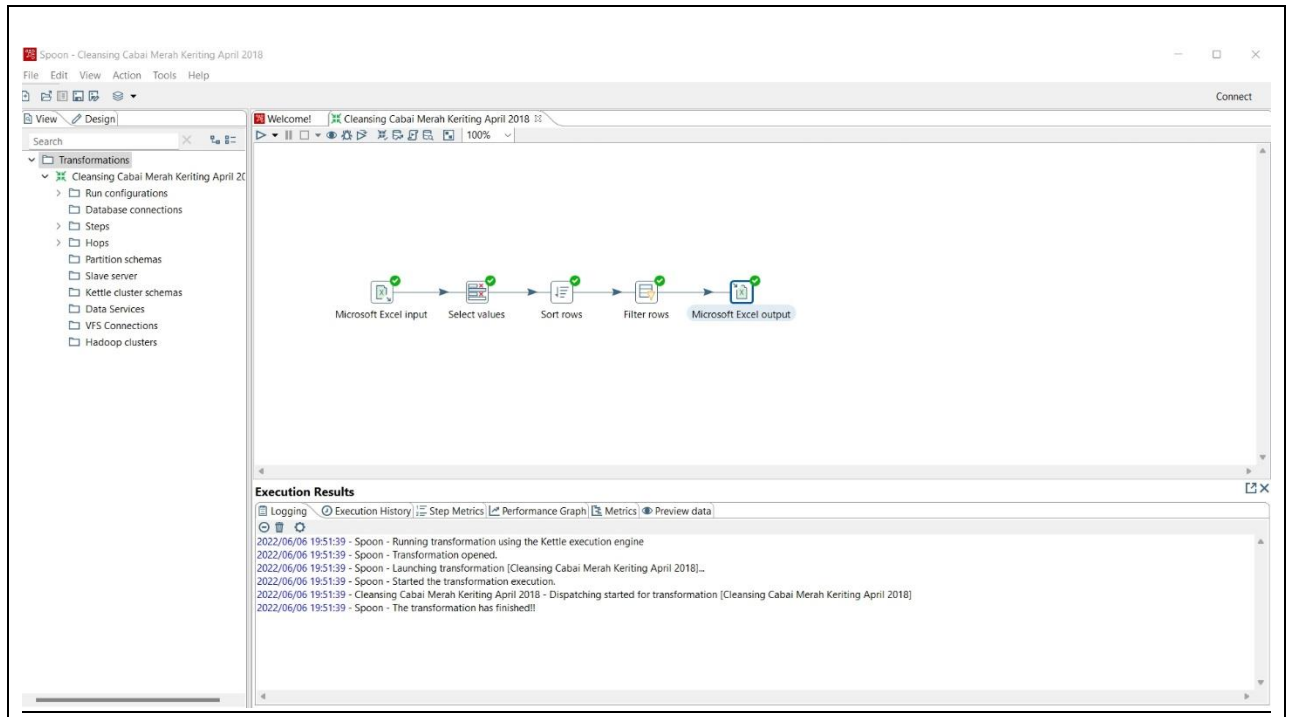


### 3. Cleansing Maret 2018

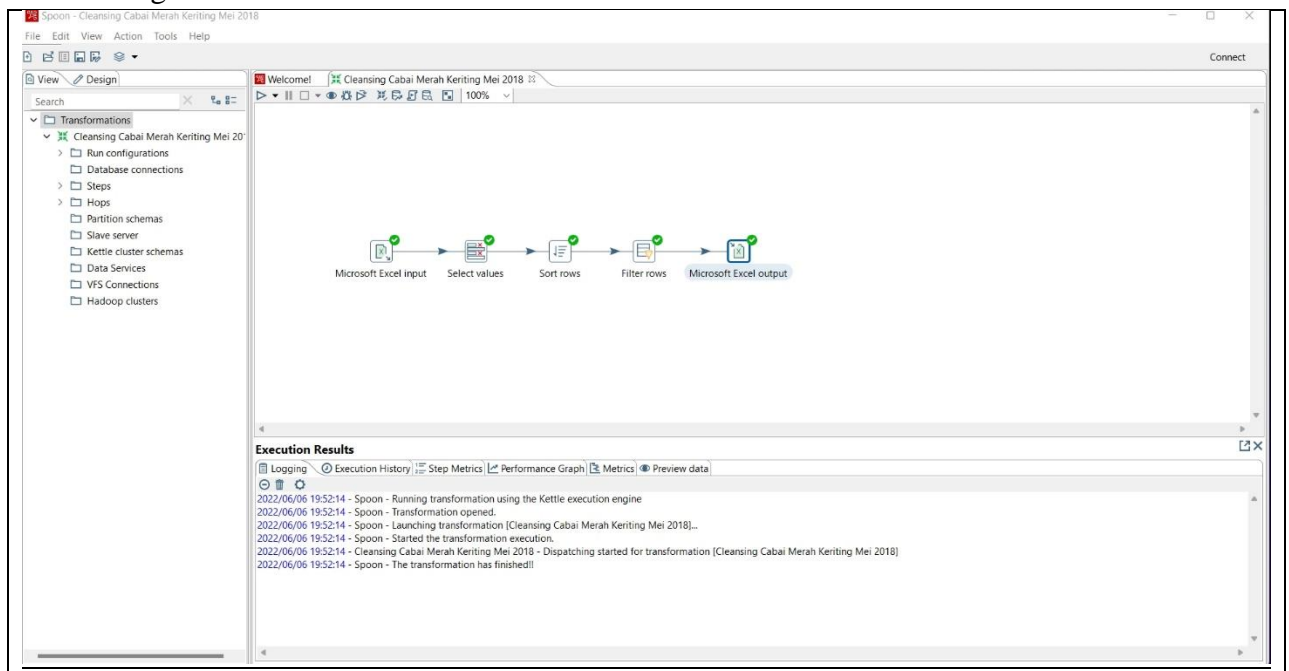


### 4. Cleansing April 2018



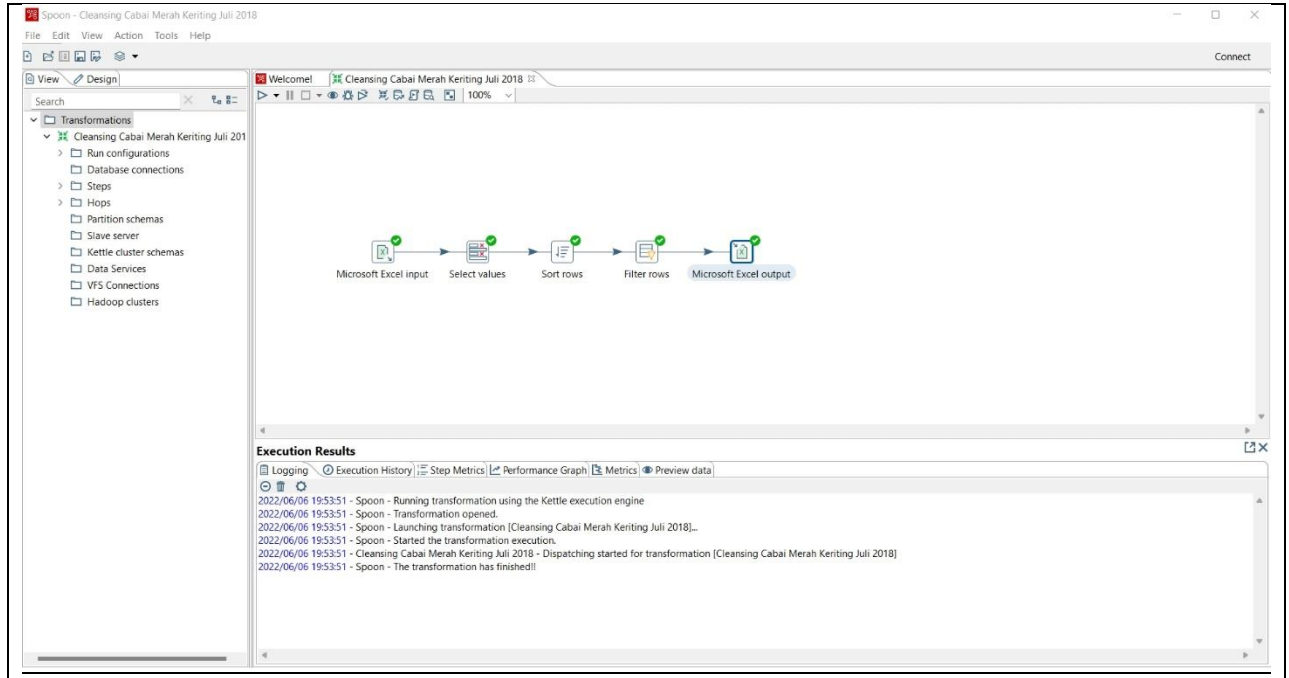


## 5. Cleansing Mei 2018

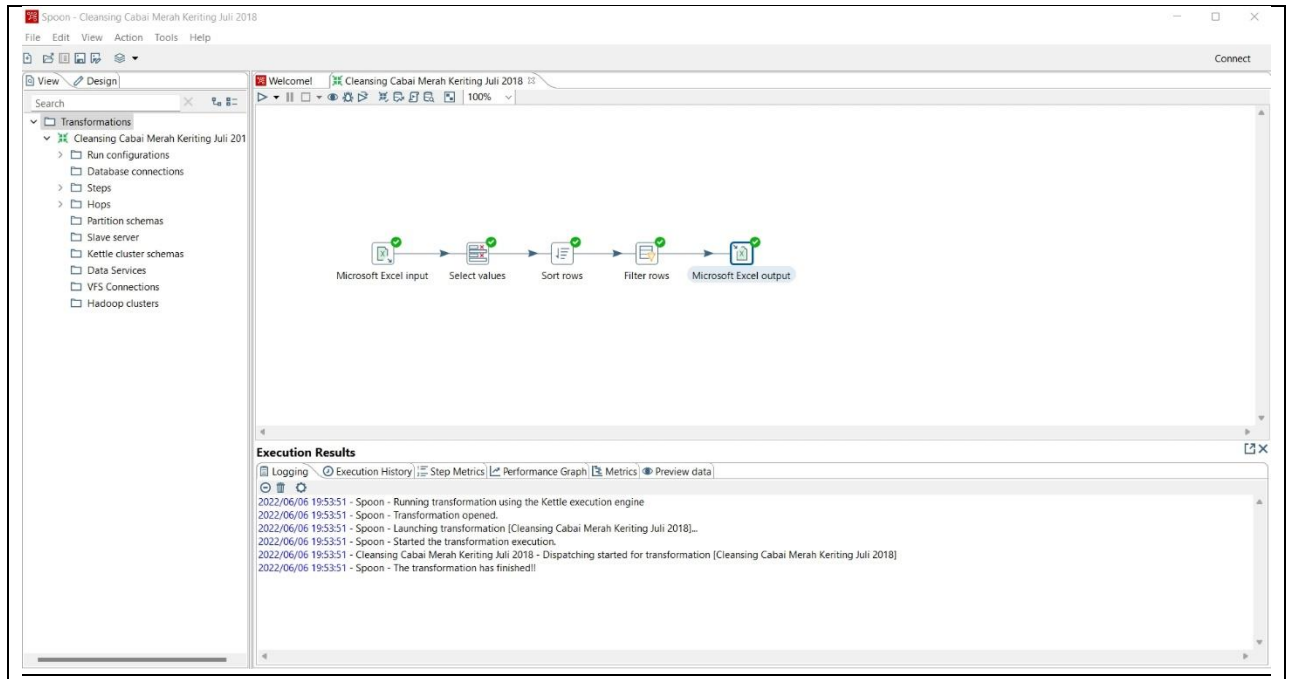


## 6. Cleansing Juni 2018



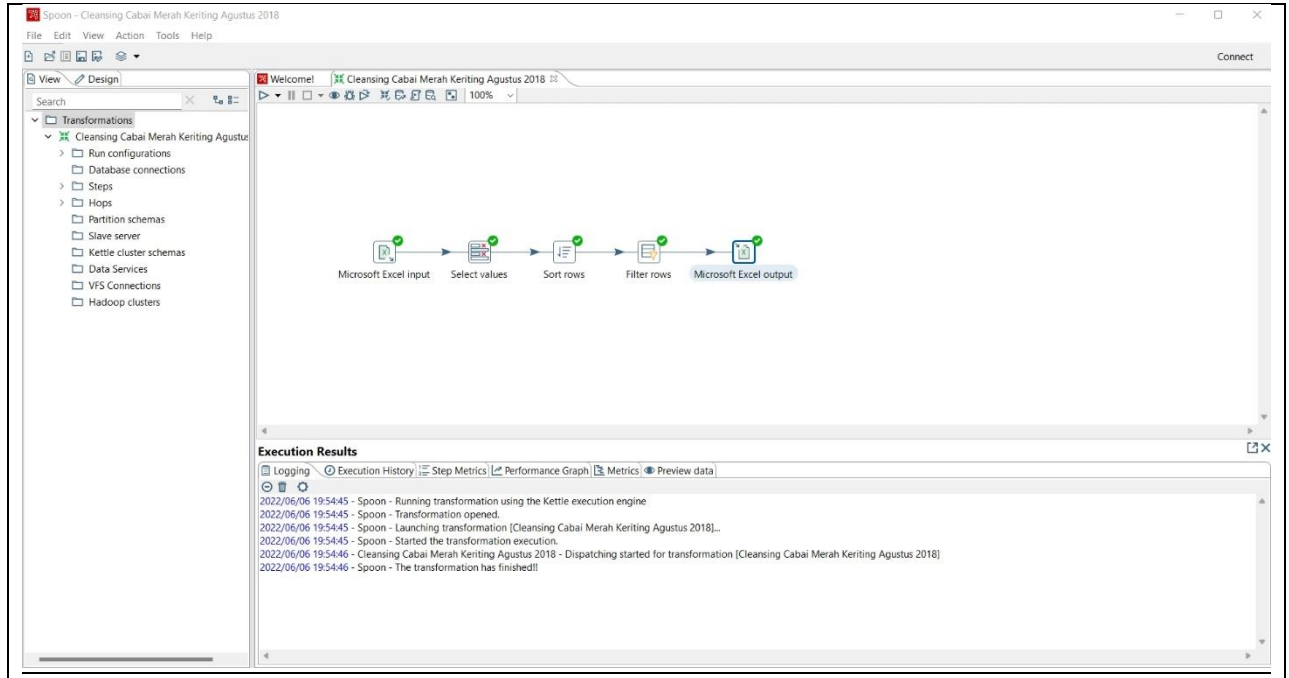


## 7. Cleansing Juli 2018

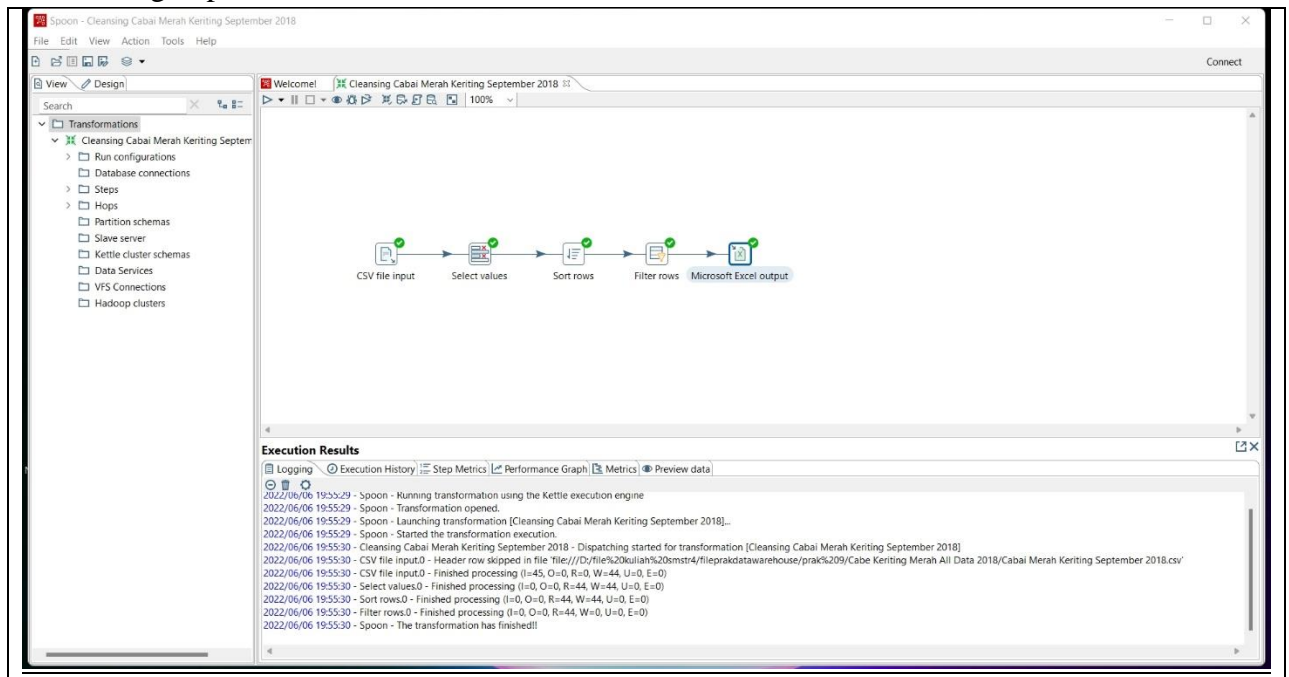


## 8. Cleansing Agustus 2018



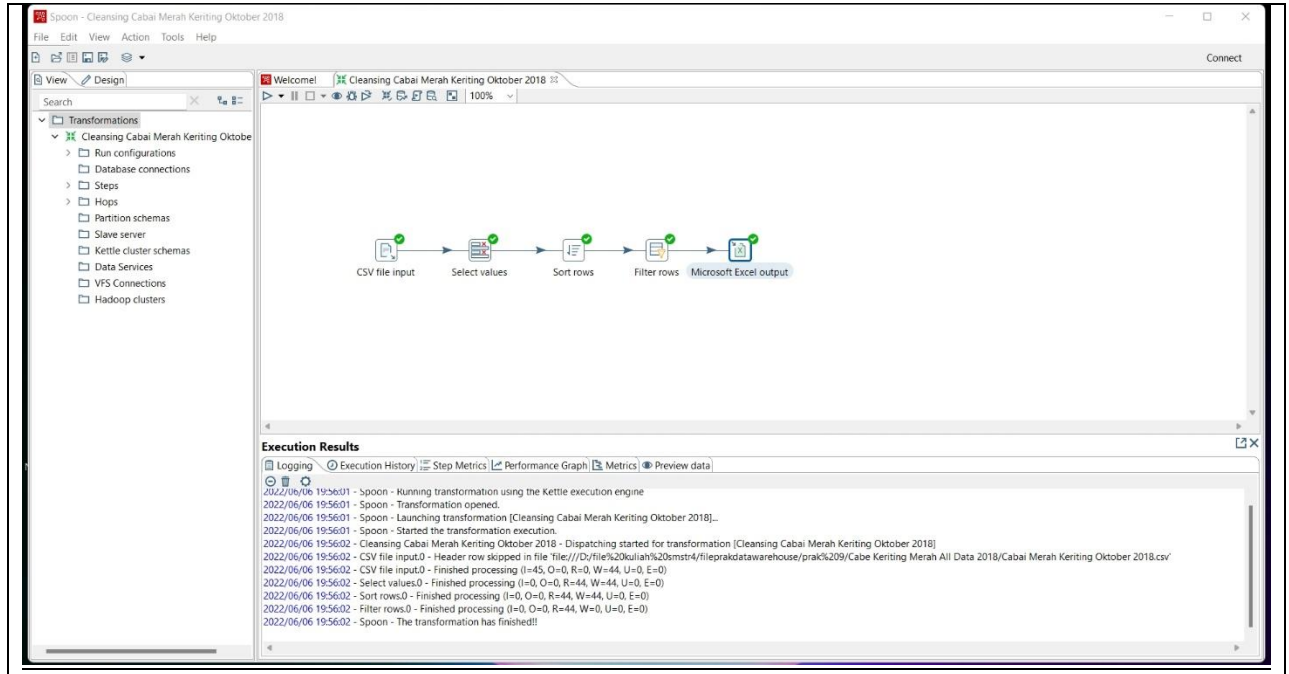


## 9. Cleansing September 2018

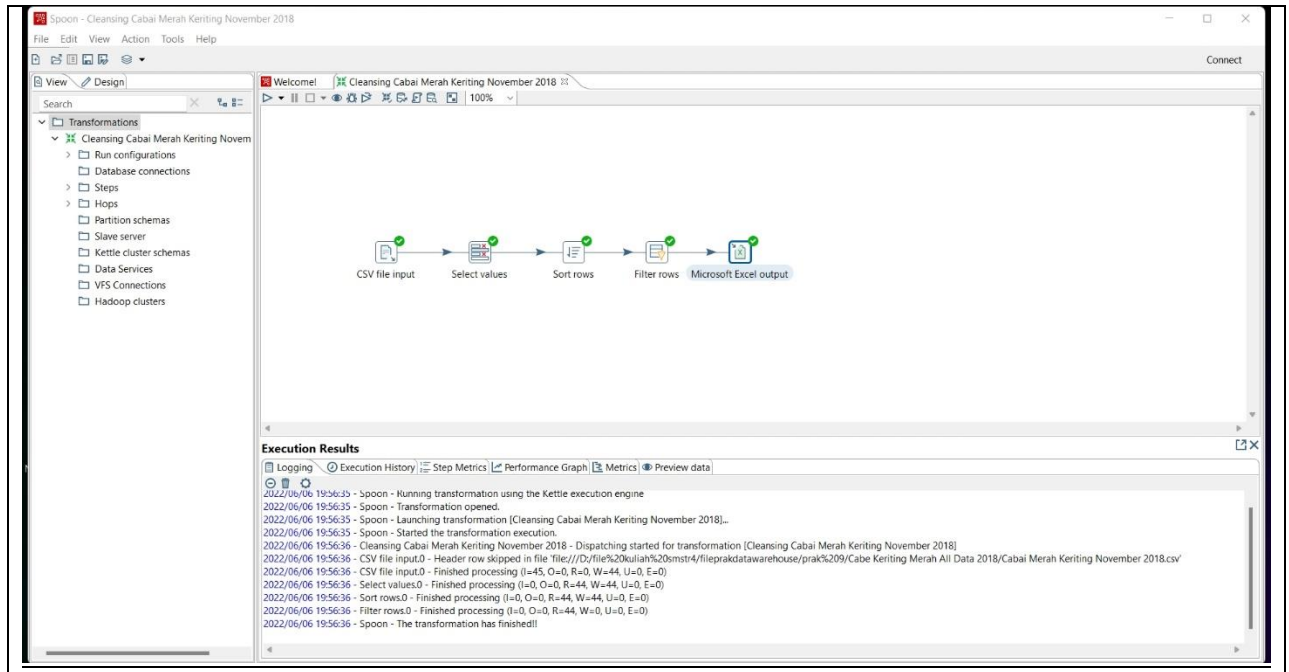


## 10. Cleansing Oktober 2018



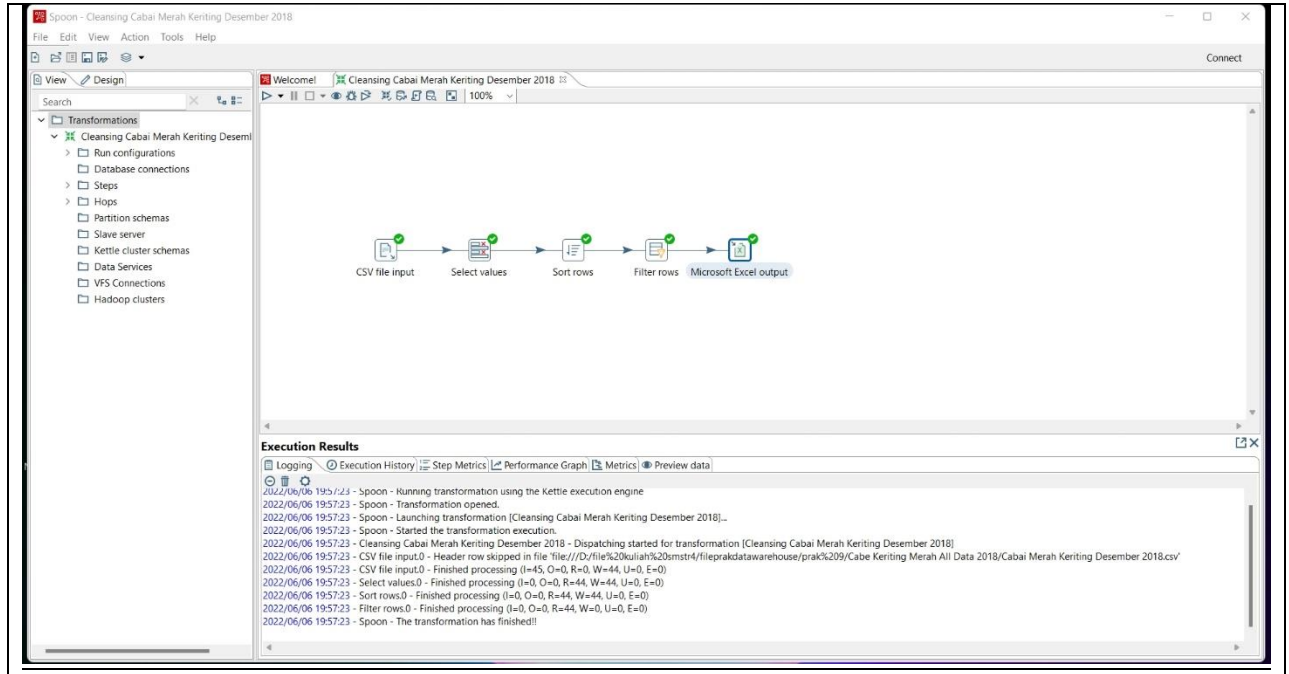


## 11. Cleansing November 2018



## 12. Cleansing Desember 2018





#### 4. File Praktikum

Github Repository:

<https://github.com/IbnuFajar7/Data-Warehouse/tree/main/Prak-9>

#### 5. Soal Latihan

Soal:

1. Apa itu Cleansing Data?
2. Mengapa Data perlu dilakukan Cleansing?

Jawaban:

1. Data cleansing adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat.
2. Menghilangkan kesalahan dan inkonsistensi yang muncul saat beberapa data sources dikumpulkan dalam satu dataset. Meningkatkan efisiensi kerja karena proses ini akan memudahkan Anda dan tim pengolah data untuk menemukan apa yang dibutuhkan dari data.

#### 6. Kesimpulan

- a. Dalam pengerjaan praktikum Data Warehouse, kita harus benar-benar teliti dalam menginputkan suatu fungsi untuk menampilkan suatu keluaran pada layar dengan sesuai.
- b. Kita dapat mengetahui bahwa kita bisa melakukan cleansing data yang bisa memudahkan kita untuk cek data yang kemungkinan kita salah input, dengan metode ini kita akan bisa lebih mudah tanpa harus cek satu-satu dan memakan waktu yang lebih lama, waktu ini akan membuat kita lebih efisien.



**7. Cek List (✓)**

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak Selesai
1.	Tugas Akhir Praktikum I	✓	

**8. Formulir Umpan Balik**

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Tugas Akhir Praktikum I	30 Menit	1

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang

